

Inter- and intra-observer variability in three grading systems for oral epithelial dysplasia

Lakshmi Krishnan, Karuppiyah Karpagaselvi¹, Jayalakshmi Kumarswamy¹, Sudheendra US², Santosh KV³, Archana Patil¹

Private Dental Practitioner, Mysuru, ¹Department of Oral Pathology and Microbiology, Vydehi Institute of Dental Sciences and Research Center, Bengaluru, ²Department of Oral Pathology and Microbiology, Coorg Institute of Dental Sciences, Virajpet, ³Department of Pathology, Vydehi Institute of Medical Sciences and Research Center, Bengaluru, India

Abstract

Background and Objectives: Management of oral premalignant lesions depends on clinical assessment and grading of oral epithelial dysplasia (OED), which remains one of the most important predictors of malignant potential. Numerous grading systems for OED exist with varying sets of assessing criteria and are largely considered subjective. The present study attempted to assess the inter- and intra-observer variability in three grading systems: Binary system, WHO (2005) and Ljubljana systems.

Methodology: Histopathological grading of 63 cases of leukoplakia was performed by two oral pathologists and one general pathologist, who were blinded, using all the three grading systems at different time intervals and was repeated twice. Inter- and intra-observer variability was then evaluated by multivariate kappa analysis.

Results: Inter-observer agreement in the two set of observations was found to be “slight” in WHO ($k = 0.001$ and 0.039), “slight” and “poor” in binary ($k = 0.108$ and -0.007), “poor” and “slight” in Ljubljana’s ($k = -0.027$ and 0.106) grading systems. Intra-observer agreement ranged from “slight” to “fair” ($k = 0.128$ and 0.295) in WHO, “fair” to “moderate” ($k = 0.224$ and 0.420) in binary and “slight” to “fair” ($k = 0.161$ and -0.353) in Ljubljana’s grading systems.

Conclusions: The binary system of classification proved to have an overall better inter- and intra-observer agreement. This study also showed better intra-observer agreement in all the grading systems as well as in individual histopathological parameters. Defining the individual parameters more objectively with the reproducible structuring of the grading systems and training of the pathologists would help reduce the variability in diagnosing dysplasia.

Key Words: Histological grading, inter and intra-observer variability, oral epithelial dysplasia, oral premalignant disorder

Address for correspondence:

Dr. Lakshmi Krishnan, #1, Behind Ridge House, 1st Stage, Gayathripuram, Mysore, Karnataka, India. E-mail: drlakshmikrishnan@gmail.com

Received: 29.12.2015, Accepted: 25.05.2016

INTRODUCTION

Oral cancer is a major health problem in many parts of the world. While its incidence is relatively low in most Western

countries; the incidence in the Indian subcontinent and in the other parts of Asia, however, remains one of the most important forms of cancer. Histologically, over 95% of oral

Access this article online	
Quick Response Code:	Website: www.jomfp.in
	DOI: 10.4103/0973-029X.185928

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Krishnan L, Karpagaselvi K, Kumarswamy J, Sudheendra US, Santosh KV, Patil A. Inter- and intra-observer variability in three grading systems for oral epithelial dysplasia. J Oral Maxillofac Pathol 2016;20:261-8.

cancers are squamous cell carcinoma (SCC), and hence, it is the most common malignant neoplasm of the oral cavity.^[1] Mortality, morbidity and cost of treatment associated with the disease increases proportionately with diagnostic delay.^[2] A vast majority of oral SCCs arise from premalignant precursors. Early detection at this stage would result in mitigation of these undesirable consequences.^[3]

The diagnosis of potentially malignant lesions is based both on clinical examination followed by an assessment of morphology and grading on histology. The histological connotation to premalignancy is marked by aberrant and uncoordinated cellular proliferation depicted basically at the cellular level (atypia), reflections of which could be discerned at tissue levels (dysplasia). It is believed that chances of malignant transformation increases with increasing severity of dysplasia thus, influencing the clinical decision-making.^[4]

Novel detection systems such as visualization methods, digital imaging and assessment of alterations in molecular and genetic characteristics have been utilized towards this end. However, none of these methodologies have been found suitable or appropriate for routine clinical use.^[4,5]

In these circumstances, histopathological grading of oral epithelial dysplasia (OED) remains one of the most important predictors of malignant potential. Most of the grading systems put forward for histopathological assessment of OED utilize multiple histologic features and scoring criteria which have long been recognized to be subjective and result in inter- and intra-personal variability.^[6,7] Any grading system is said to be clinically useful if it is reproducible with no inter- and intra-observer variability.^[8]

The WHO classification (2005) is frequently used to grade OED and includes hyperplasia, three grades of dysplasia, namely mild, moderate, severe and carcinoma-*in situ* (CIS).^[9] The Ljubljana classification, proposed for grading epithelial hyperplastic laryngeal lesions, has been applied for grading of oral epithelial dysplastic lesions. This system recognizes four categories: Simple and abnormal hyperplasia both being benign, atypical hyperplasia as premalignant and CIS.^[10] The new binary grading system (2006) has shown encouraging predictive value in detecting malignant transformation in OED by eliminating “opt-out” judgements on cases that are diagnosed with either the four-scale or the five-scale grading system.^[11]

This study aims to evaluate the reproducibility of the binary system of grading OED and compare the inter- and intra-observer variability in grading of OED in a binary system with that of WHO and Ljubljana system of classifications.

METHODOLOGY

The study was conducted on hematoxylin and eosin (H and E) stained sections obtained from the biopsied tissue specimens and graded for OED using three classification systems: WHO (2005), Ljubljana and binary systems of classifications.

The WHO classification (2005) is based on a combination of seven architectural and nine cytological changes [Table 1] with more explicit consideration for levels of change within the epithelium.^[9]

The WHO classification:^[9]

- **Hyperplasia:** This describes increased cell numbers. This may be in the spinous layer (acanthosis) and/or in the basal/parabasal cell layers, termed basal cell hyperplasia. The architecture shows regular stratification without cellular atypia.

When an architectural disturbance is accompanied by cytologic atypia, the term dysplasia applies. Dysplasia is a spectrum and no criteria exist to precisely divide this spectrum into mild, moderate and severe categories.

- **Mild dysplasia:** Architectural disturbance limited to the lower third of the epithelium accompanied by cytological atypia
- **Moderate dysplasia:** Architectural disturbance extending into the middle-third of the epithelium. However, consideration of the degree of cytologic atypia may require upgrading
- **Severe dysplasia:** Greater than two-third of the epithelium showing architectural disturbance with associated cytologic atypia
- **CIS:** Full thickness or almost full thickness architectural abnormalities in the viable cellular layers accompanied by pronounced cytologic atypia. Atypical mitotic figures and abnormal superficial mitoses are commonly seen in CIS.

Table 1: WHO architectural and cytological criteria to classify OED

Architecture criteria	Cytology criteria
Irregular epithelial stratification	Abnormal variation in nuclear size
Loss of polarity of basal cells	Abnormal variation in nuclear shape
Drop-shaped rete ridges	Abnormal variation in cell size
Increased number of mitotic figures	Abnormal variation in cell shape
Abnormally superficial mitoses	Increased nuclear/cytoplasmic ratio
Premalignant keratinization in single cells	Increased nuclear size
Keratin pearls within rete ridges	Atypical mitotic figures
	Increased number and size of nucleoli
	Hyperchromatism

The Ljubljana classification:^[10]

- Simple hyperplasia: A benign hyperplastic process with retention of the normal pattern of epithelium, which is thickened because of increased prickle cell layer. The cellular components of basal and parabasal regions remain unchanged. There is no cellular atypia
- Abnormal hyperplasia: A benign augmentation of basal and parabasal layers. They are augmented to a degree which constitutes up to one-half of the total epithelial thickness. Stratification is fully retained. Occasionally, more than this portion of the epithelium may be involved by the hyperplastic cells without significant atypical nuclear changes. Nuclei in the cells of the augmented basal and parabasal layers may be moderately enlarged but still maintain a uniform distribution of nuclear chromatin. Occasional typical mitoses may be found in or near basal layer. Small numbers of epithelial cells, <5%, are dyskeratotic
- Atypical hyperplasia: Risky epithelium demonstrating a recognizable alteration of epithelial cells toward malignancy, but not to such a degree as is seen in carcinomatous cells. Stratification is still preserved in the general epithelial structure. The nuclei are enlarged and nuclear contour may be irregular with marked variation in staining intensity. The nuclear/cytoplasmic ratio is increased. Mitotic figures are increased, but not numerous and they are found within two-third of the epithelium above the basement membrane. They are rarely, if ever, abnormal. Dyskeratotic cells are frequent. Civatte bodies (apoptotic cells) may be present
- CIS: Shows features of carcinoma without invasion. Stratification of the epithelium as a whole is lost. Marked cellular alteration of the type found in atypical hyperplasia is present to a considerably greater degree. Many mitotic figures are present throughout the epithelium, including its upper one-third; and abnormal mitoses are frequently found.

The binary system (2006)^[11] uses the same WHO architectural and cytological criteria to categorize OED into:

- “Low risk” (<4 architectural changes and <5 cytological changes) and
- “High risk” (at least four architectural changes and five cytological changes).

Study design

Sixty-three cases obtained from buccal mucosa of as many individuals which were clinically diagnosed as leukoplakia and histopathologically diagnosed as OED formed the study sample. These did not include cases of candidal leukoplakia, proliferative verrucous leukoplakia and cases adjacent to or in association with a previous diagnosis of oral SCC. The original

sign-out diagnosis was done according to WHO classification and consisted of 16 mild, 34 moderate and 13 severe dysplasia cases. Three independent observers which included two oral pathologists and one general pathologist evaluated the H and E slides. They were blinded to the initial diagnosis, and the clinical data and score sheets were made for all the three grading system to ensure standardization of reporting. The slides were assessed on different day intervals for each grading systems and the process was repeated again after 3 months at which time they were renumbered to eliminate bias for the second round of grading. Furthermore, individual features such as irregular epithelial stratification, loss of polarity of basal cells, drop-shaped rete ridges, increased number of mitotic figures, abnormally superficial mitosis, premature keratinization in single cells, keratin pearls within rete ridges, abnormal variation in nuclear size and shape, cell size and shape, increased nuclear/cytoplasmic ratio and nuclear size, atypical mitotic figures, increased number and size of nucleoli and hyperchromatism were analyzed.

Statistical analysis

The data collected was analyzed using statistical software SPSS (SPSS Inc. Released 2007. SPSS for Windows, Version 16.0. Chicago). Kappa statistics was used for assessment of both inter- and intra-observer agreement. Value of *k* was considered as <0.00: Poor; 0.00–0.20: Slight; 0.21–0.40: Fair; 0.41–0.60: Moderate; 0.61–0.80: Good.^[12,13] *P* < 0.05 were considered to be statistically significant.

RESULTS

Inter-observer agreement

Inter-observer variability was assessed between the three observers for all three grading systems during both the observations. The kappa values, its strengths of agreement, percentage and probability values of inter-observer observations are tabulated in Table 2.

- WHO system: During the first observation, out of the 63 cases graded using WHO grading system, the first observer graded 7, 25 and 31 cases as mild dysplasia, moderate dysplasia and severe dysplasia respectively; the second observer graded 8, 29 and 26 cases as mild dysplasia, moderate dysplasia and severe dysplasia respectively; and the third observer graded 28, 34 and 1 cases as mild dysplasia, moderate dysplasia and severe dysplasia respectively. The agreement between observers 1 and 2 showed a kappa score of 0.021 (slight agreement), between 1 and 3 the score was 0.020 (slight agreement) and between 2 and 3 was –0.037 (poor agreement). The results of the statistical analysis between all the investigators indicated the absence of a significant difference. During the second observation, the kappa scores for agreement between observers 1 and 2, 1 and 3 and 2 and 3 were 0.129 (slight

- agreement), -0.034 (poor agreement) and 0.024 (slight agreement), respectively
- b. Binary system: Out of the 63 cases graded using binary grading system, the first observer graded 10 and 53 cases as low- and high-risk lesions, respectively; the second observer graded 22 and 41 cases as low- and high-risk lesions, respectively; and the third observer graded 50 and 13 cases as low- and high-risk lesions respectively with kappa scores ranging between slight agreement to fair agreement. During the second observation, the kappa scores for agreement between the oral pathologists was 0.147 (poor agreement)
 - c. Ljubljana system: Out of the 63 cases graded using Ljubljana grading system, the first observer graded 1, 29 and 33 cases as abnormal hyperplasia, atypical hyperplasia and CIS lesions, respectively; the second observer graded 10, 34 and 19 cases as abnormal hyperplasia, atypical hyperplasia and CIS lesions, respectively; and the third observer graded 31, 31 and 1 cases as abnormal hyperplasia, atypical hyperplasia and CIS lesions, respectively, with slight agreement between observers 1 and 2; and 2 and 3. There was poor agreement between observers 1 and 3 with kappa score of -0.090. In the second observation, kappa scores ranged from poor to fair agreement.

Intra-observer agreement

The kappa values with its strength of agreement and percentage for all intra-observer observations along with probability values are shown in Table 3.

- a. WHO system: Absolute agreement for observer 1, 2 and 3 between first and second observation was 58.73%, 46.03% and 60.31%, respectively. The kappa scores for observers 1 and 3 were “fair” and statistically significant
- b. Binary system: Absolute agreement was relatively better in this system. For observers 1, 2 and 3, absolute agreement between first and second observation was 80.95%, 65.07% and 79.36% with a “fair” kappa score for observers 1 and 2 (not statistically significant) and a statistically significant “moderate” kappa score for observer 3
- c. Ljubljana system: For observers 1, 2 and 3, absolute agreement between first and second observation was 58.73%, 46.03% and 66.66% with a kappa score of “fair” and “slight” for observers 1 and 2 (not statistically significant), respectively and a statistically significant “fair” kappa score for observer 3.

Agreements between the observers on each of the 16 individual parameters assessed according to the binary system of classification are presented in Table 4.

DISCUSSION

The subjectivity in assessing OED has often been raised which is largely due to lack of well-defined criteria that can be used as guidelines for grading.^[8] Although most oral pathologists possibly recognize and accept the criteria for grading epithelial dysplasia, there is great variability in their interpretation of the

Table 2: Kappa values, its strength of agreement, percentage and probability values for all inter-observer observations

Observer pairs		First observation			Second observation		
		K	P	Percentage	K	P	Percentage
First observer versus second observer	WHO	0.021 (slight)	0.826	41.26	0.129 (slight)	0.147	46.03
First observer versus third observer		0.020 (slight)	0.735	28.57	-0.034 (poor)	0.473	19.04
Second observer versus third observer		-0.037 (poor)	0.564	28.57	0.024 (slight)	0.632	22.22
First observer versus second observer	Binary	0.201 (fair)	0.070	68.25	-0.147 (poor)	0.155	55.55
First observer versus third observer		0.094 (slight)	0.079	36.50	0.094 (slight)	0.077	38.09
Second observer versus third observer		0.029 (slight)	0.725	42.85	0.032 (slight)	0.721	44.44
First observer versus second observer	Ljubljana	0.006 (slight)	0.946	41.26	0.333 (fair)	0.000	58.73
First observer versus third observer		-0.090 (poor)	0.052	17.46	-0.051 (poor)	0.335	25.39
Second observer versus third observer		0.001 (slight)	0.987	34.92	0.038 (slight)	0.581	33.33

P<0.05 were considered to be statistically significant

Table 3: Kappa values with its strength of agreement and percentage for all intra-observer observations along with probability values

Grading systems	Values	First observer	Second observer	Third observer
		First observation versus second observation	First observation versus second observation	First observation versus second observation
WHO	K	0.029 (fair)	0.128 (slight)	0.256 (fair)
	P	0.003	0.159	0.016
	Percentage	58.73	46.03	60.31
Binary	K	0.224 (fair)	0.232 (fair)	0.420 (moderate)
	P	0.073	0.066	0.001
	Percentage	80.95	65.07	79.36
Ljubljana	K	0.221 (fair)	0.161 (slight)	0.353 (fair)
	P	0.055	0.068	0.003
	Percentage	58.73	46.03	66.66

P<0.05 were considered to be statistically significant

Table 4: Observations of the individual parameters “present” according to the binary system of classification (first set)

Histological criteria/ parameters (P1-P16)	First observer		Second observer		Third observer		Total agreement (%)
	Yes	No	Yes	No	Yes	No	
<i>P1</i>							
<i>n</i>	63	0	63	0	63	0	100
Percentage	100	0	100	0	100	0	
<i>P2</i>							
<i>n</i>	61	2	58	5	62	1	92.05
Percentage	96.82	3.17	92.06	7.93	98.41	1.58	
<i>P3</i>							
<i>n</i>	31	32	25	38	28	35	76.18
Percentage	49.20	50.79	39.68	60.31	44.44	55.55	
<i>P4</i>							
<i>n</i>	53	10	50	13	22	41	59.72
Percentage	84.12	15.87	79.36	20.63	34.92	65.07	
<i>P5</i>							
<i>n</i>	48	15	27	36	1	62	47.61
Percentage	76.19	23.8	42.85	57.14	1.58	98.41	
<i>P6</i>							
<i>n</i>	28	35	26	37	6	57	49.99
Percentage	44.44	55.55	41.26	58.73	9.52	90.47	
<i>P7</i>							
<i>n</i>	13	50	5	58	1	62	66.66
Percentage	20.63	79.36	7.93	92.06	1.58	98.41	
<i>P8</i>							
<i>n</i>	63	0	63	0	63	0	100
Percentage	100	0	100	0	100	0	
<i>P9</i>							
<i>n</i>	63	0	63	0	63	0	100
Percentage	100	0	100	0	100	0	
<i>P10</i>							
<i>n</i>	63	0	55	8	63	0	87.3
Percentage	100	0	87.3	12.69	100	0	
<i>P11</i>							
<i>n</i>	63	0	59	4	62	1	92.06
Percentage	100	0	93.65	6.34	98.41	1.58	
<i>P12</i>							
<i>n</i>	63	0	63	0	63	0	100
Percentage	100	0	100	0	100	0	
<i>P13</i>							
<i>n</i>	63	0	62	1	63	0	98.4
Percentage	100	0	98.41	1.58	100	0	
<i>P14</i>							
<i>n</i>	37	26	30	33	0	63	47.61
Percentage	58.73	41.26	47.61	52.38	0	100	
<i>P15</i>							
<i>n</i>	35	28	55	8	35	28	61.05
Percentage	55.55	44.44	87.30	12.69	55.55	44.44	
<i>P16</i>							
<i>n</i>	52	11	54	9	42	21	72.21

P1: Irregular epithelial stratification, *P2*: Loss of polarity of basal cells, *P3*: Drop shaped rete ridges, *P4*: Increased number of mitotic figures, *P5*: Abnormally superficial mitosis, *P6*: Premature keratinisation in single cells, *P7*: Keratin pearls within rete ridges, *P8*: Abnormal variation in nuclear size, *P9*: Abnormal variation in nuclear shape, *P10*: Abnormal variation in cell size, *P11*: Abnormal variation in cell shape, *P12*: Increased nuclear/cytoplasmic ratio, *P13*: Increased nuclear size, *P14*: Atypical mitotic figures, *P15*: Increased number and size of nucleoli, *P16*: Hyperchromatism, n: number

presence, degree and significance of the individual criteria.^[14,15] Grading is hampered by the arbitrary division into distinct categories of a continually progressing process without naturally and sharply defined borders. Grading is, in fact, an attempt to impose discrete categories on what is in effect a continuous grey scale and therefore, any grading scheme is by definition artificial. The ultimate diagnosis depends on the emphasis which is put on each of these categories in grading by observers and is, therefore, subjective.^[11]

Most of the previous studies have compared the inter-observer variability between either oral pathologists or between general pathologists. In only two studies, comparisons were made between the observations of oral and general pathologists in grading OED.^[11,16] In our study, two oral and one general pathologists performed the evaluation of OED. Many of the studies carried out to assess the inter-observer and intra-observer variability in epithelial dysplasia have compared not more than two systems of classification in contrast to our

study where three systems of classifications of OED have been compared. To the best of our knowledge, ours is the first study which evaluated binary system in comparison with both WHO (2005) and Ljubljana system of classifications for grading OED.

The inter-observer reliability of our study in WHO (2005) grading system showed “slight” agreement ($k = 0.001$, 32.8%; $k = 0.039$, 29.09%) both times, despite the less objective criteria and its arbitrary division of epithelium into three-third. This is probably because WHO system is considered less cumbersome and is more commonly used. Similar to our results, in a previous study for six examiners, the average agreement between the observers was 38.5% ($k = 0.17$).^[17] In a study using a five-point ordinal scale, the group showed “fair” agreement ($k = 0.37$)^[18] and in another study using WHO classification 2005, kappa values obtained ranged from 0.06 to 0.43 (slight to fair agreement).^[11] The intra-observer agreement in WHO grading system in our study was found to be “fair” ($k = 0.295$, first observer; $k = 0.256$ third observer) and “slight” ($k = 0.128$, second observer). Kappa values from the previous studies, varied from 0.30–0.83^[18] to 0.05–0.49.^[17]

The inter-observer variability in our study in the Ljubljana grading system showed “poor” ($k = -0.027$) and “slight” agreement ($k = 0.106$) in the first and second observations respectively; in contrast to “good” agreement ($k = 0.52$) in previous studies done in laryngeal dysplasias despite the criteria being clearly defined.^[10] As this grading system is not commonly used, experience in using it in OED should be taken into consideration.^[19] The usage of this system may be encouraged due to its more definitive criteria. In Ljubljana’s grading system, the intra-observer agreement in our study was “fair” ($k = 0.221$ first observer; $k = 0.353$ third observer) and “slight” ($k = 0.161$ second observer).

In binary grading system (based on the WHO 2005 classification), there was “slight” inter-observer agreement ($k = 0.108$) in the first observation and “poor” agreement ($k = -0.007$) in the second observation in contrast to the “moderate” agreement ($k = 0.50$) in a previous study done in 2006,^[11] despite the criteria being clearly defined. This variation might be related to the difference in understanding of the criteria/features.^[15] With the binary grading system, the intra-observer agreement in our study was found to be “fair” ($k = 0.224$ first observer; $k = 0.232$ second observer) and “moderate” ($k = 0.420$ third observer). The intra-observer agreement was better in binary system (29.2%) than in the other two grading systems (Ljubljana system – 24.5% and WHO system – 22.63%).

Comparison of inter- and intra-observer agreement analyses in the present study showed that the intra-observer reproducibility

levels surpassed those of inter-observer agreement indicating that observer bias is present. Different pathologists seem to have their own interpretations and opinions of histopathological definitions of epithelial dysplastic lesions.^[18]

The binary system adopts a two-point scoring system as opposed to the multiple scoring system used in the other grading systems. Although this may ease in the categorization of disease and possibly decrease observer variability, the prognostic value regarding clinical outcome is largely untested.

Although the grading systems vary in the method of scoring, they involve a similar utilization of a compendium of individual morphologic characteristics. Inconsistency in the construal of individual features of epithelial dysplasia could significantly contribute to the inter-observer variability.

Although the WHO (2005) classification defined each category and pointed out how to grade the OED lesions, no definitions or exemplar photomicrographs that could illustrate the individual architectural and cytological features were included. In the absence of such information, assessment of these features would depend mainly on the educational process of grading OED.^[15] In a previous study using Smith and Pindborg method, features which predominated in severely dysplastic lesions were, mitotic activity superficial to basal layer (98%), increased nuclear/cytoplasmic ratio (85%), increased mitotic activity (84%), pleomorphic cells and nuclei (79%), irregular epithelial stratification (75%), basal cell hyperplasia (73%), loss of polarity (72%) and loss of cellular adherence (71%).^[20] In a similar study, the individual histopathological features were evaluated to predict the malignant transformation among four pathologists using WHO 2005 classification. There was highest agreement among all the four pathologists in increased number of mitosis, drop shaped rete ridges, increased nuclear/cytoplasmic ratio and cellular pleomorphism. The highest disagreement was noted in irregular epithelial stratification, loss of polarity, nuclear pleomorphism, abnormal mitosis and hyperchromatism.^[15]

In this study, absolute agreement among all three observers was seen in the assessment of irregular epithelial stratification, abnormal variation in nuclear size and nuclear shape and increased nuclear/cytoplasmic ratio. Considerable agreement was also seen in the following features (in the decreasing order): Increased nuclear size (98.4%), loss of polarity (92.5%), abnormal variation in cell shape (92.06%), abnormal variation in cell size (87.30%), drop shaped rete ridges (76.18%), hyperchromatism (72.21%), presence of keratin pearls in rete ridges (66.66%) and increased number and size of nucleoli (61.05%). Highest disagreement was seen in abnormally superficial mitosis (47.61%), atypical mitosis (47.61%), premature keratinization (49.99%) and

increased mitosis (59.72%). These four features found considerable agreement between the oral pathologists and the disagreement existed between the general pathologist and oral pathologists participating in our study. The general pathologist (observer 3) showed less agreement with the oral pathologist, as also observed in another study.^[11] The reason for this could be that the dysplasia to carcinoma sequence theory was introduced from the viewpoint of pathologic changes in the uterine cervix; in contrast, almost all premalignant lesions of the oral mucosa show superficial maturation and differentiation.^[21]

In a study, irregular epithelial stratification and presence of drop shaped rete pegs were the most frequent manifestations of dysplasia.^[22] In this study, all three observers observed the presence of irregular epithelial stratification in all the slides used in our study, irrespective of the severity of dysplasia. Whereas, drop shaped rete ridges were present only in 14 of the 31 severe dysplastic cases (observer 1), 13 out of the 28 severe dysplastic cases (observer 2) and 1 out of the only 1 severe dysplasia case (observer 3).

In a study done in 1995, the inter-observer agreement between the observers was 38.5% ($k = 0.17$), which compared to those from the previous study, where the same examiners had evaluated same slides without clinical histories, represented a 2.5–20% decrease in the inter-observer agreement.^[23] Furthermore, patients with dysplasias in more than one site had a slightly higher probability of being diagnosed as either moderate/severe.^[24] Clinical details of site of lesion, the age of the patient and associated habits may help in removal of subjectivity and thus improve agreement between observers.

The subjectivity in the evaluation of the established criteria of grading; arbitrary division of the grading; lack of calibration of the used criteria and grading; and the lack of sufficient knowledge of which criteria are more important for the prediction of malignant potential are attributed for the lack of agreement on grading oral dysplasia lesions.^[8,25] Certain flaws, advantages and biases, however, are inherent in investigations that use multiple examiners.^[26]

So, improvement in the standard of the histopathology reporting of oral epithelial dysplastic lesions could be achieved by consideration of several points. Of these, paramount is a need for universal definition of the individual parameters (architectural and cytological features) that are the basis of any OED grading process.

CONCLUSIONS

Morphologic assessment of epithelial dysplasia has traditionally been used as an indicator of malignant transformation.

Advanced tools like an assessment of molecular or genomic alterations may not be feasible in all situations. The consistency of the grading epithelial dysplasia with clearly delineated individual characteristics and scoring system will indeed assist in the better clinical management of the lesion.

The binary system of classification proved to have an overall better inter- and intra-observer agreement in this study, as compared to the WHO and Ljubljana system of classification of OED. However, dysplasia classification can best be improved upon by understanding the fundamental biology of the process, wherein the clinicopathological correlation may improve the inter- and intra-observer agreement. With the increasing demands for “evidence-based medicine,” subjective histopathological approaches do need to be refined; and concepts, as well as diagnostic criteria, scientifically validated.^[27]

Acknowledgment

Our sincere thanks to Dr. T. V. Narayan, ex-Head, Department of Oral Pathology and Microbiology, The Oxford Dental College and Hospital, Bengaluru; Dr. Radhika M. Bavale, Professor and Head, Department of Oral Pathology and Microbiology, Krishnadevaraya College of Dental Sciences, Bengaluru; Dr. Usha Hegde, Professor and Head, Department of Oral Pathology and Microbiology, JSS Dental College and Hospital, Mysore, for permitting us to utilize the tissue slides from their department for the study; and lab technician Mrs. Dorothy Anita, Department of Oral Pathology and Microbiology, Vydehi Institute of Dental Sciences and Research Centre.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. van Monsjou HS, Balm AJ, van den Brekel MM, Wreesmann VB. Oropharyngeal squamous cell carcinoma: A unique disease on the rise? *Oral Oncol* 2010;46:780-5.
2. Gómez I, Seoane J, Varela-Centelles P, Diz P, Takkouche B. Is diagnostic delay related to advanced-stage oral cancer? A meta-analysis. *Eur J Oral Sci* 2009;117:541-6.
3. Epstein JB, Zhang L, Rosin M. Advances in the diagnosis of oral premalignant and malignant lesions. *J Can Dent Assoc* 2002;68:617-21.
4. Bouquot JE, Speight PM, Farthing PM. Epithelial dysplasia of the oral mucosa – Diagnostic problems and prognostic features. *Curr Diagn Pathol* 2006;12:11-21.
5. Poh CF, MacAulay CE, Laronde DM, Williams PM, Zhang L, Rosin MP. Squamous cell carcinoma and precursor lesions: Diagnosis and screening in a technical era. *Periodontol* 2000 2011;57:73-88.
6. Speight PM. Update on oral epithelial dysplasia and progression to cancer. *Head Neck Pathol* 2007;1:61-6.
7. Pindborg JJ, Reibel J, Holmstrup P. Subjectivity in evaluating oral

- epithelial dysplasia, carcinoma *in situ* and initial carcinoma. *J Oral Pathol* 1985;14:698-708.
8. Warnakulasuriya S. Histological grading of oral epithelial dysplasia: Revisited. *J Pathol* 2001;194:294-7.
 9. Barnes L, Eveson JW, Reichart P, Sidransky D. World Health Organization Classification of Tumours: Pathology and Genetics of Head and Neck Tumours. Lyon: IARC Press; 2005.
 10. Zerdoner D. The Ljubljana classification – Its application to grading oral epithelial hyperplasia. *J Craniomaxillofac Surg* 2003;31:75-9.
 11. Kujan O, Oliver RJ, Khattab A, Roberts SA, Thakker N, Sloan P. Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. *Oral Oncol* 2006;42:987-93.
 12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
 13. Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic. *Fam Med* 2005;37:360-3.
 14. Warnakulasuriya S, Reibel J, Bouquot J, Dabelsteen E. Oral epithelial dysplasia classification systems: Predictive value, utility, weaknesses and scope for improvement. *J Oral Pathol Med* 2008;37:127-33.
 15. Kujan O, Khattab A, Oliver RJ, Roberts SA, Thakker N, Sloan P. Why oral histopathology suffers inter-observer variability on grading oral epithelial dysplasia: An attempt to understand the sources of variation. *Oral Oncol* 2007;43:224-31.
 16. Karabulut A, Reibel J, Therkildsen MH, Praetorius F, Nielsen HW, Dabelsteen E. Observer variability in the histologic assessment of oral premalignant lesions. *J Oral Pathol Med* 1995;24:198-200.
 17. Abbey LM, Kaugars GE, Gunsolley JC, Burns JC, Page DG, Svirsky JA, *et al.* The effect of clinical information on the histopathologic diagnosis of oral epithelial dysplasia. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 1998;85:74-7.
 18. Brothwell DJ, Lewis DW, Bradley G, Leong I, Jordan RC, Mock D, *et al.* Observer agreement in the grading of oral epithelial dysplasia. *Community Dent Oral Epidemiol* 2003;31:300-5.
 19. Michaels L. The Kambic-Gale method of assessment of epithelial hyperplastic lesions of the larynx in comparison with the dysplasia grade method. *Acta Otolaryngol Suppl* 1997;527:17-20.
 20. Schepman KP, van der Meij EH, Smeele LE, van der Waal I. Malignant transformation of oral leukoplakia: A follow-up study of a hospital-based population of 166 patients with oral leukoplakia from The Netherlands. *Oral Oncol* 1998;34:270-5.
 21. Izumo T. Oral premalignant lesions: From the pathological viewpoint. *Int J Clin Oncol* 2011;16:15-26.
 22. Bánóczy J, Csiba A. Occurrence of epithelial dysplasia in oral leukoplakia. Analysis and follow-up study of 12 cases. *Oral Surg Oral Med Oral Pathol* 1976;42:766-74.
 23. Abbey LM, Kaugars GE, Gunsolley JC, Burns JC, Page DG, Svirsky JA, *et al.* Intraexaminer and interexaminer reliability in the diagnosis of oral epithelial dysplasia. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 1995;80:188-91.
 24. Silverman S Jr., Gorsky M, Lozada F. Oral leukoplakia and malignant transformation. A follow-up study of 257 patients. *Cancer* 1984;53:563-8.
 25. Reibel J. Prognosis of oral pre-malignant lesions: Significance of clinical, histopathological, and molecular biological characteristics. *Crit Rev Oral Biol Med* 2003;14:47-62.
 26. Bouquot JE, Gorlin RJ. Leukoplakia, lichen planus, and other oral keratoses in 23,616 white Americans over the age of 35 years. *Oral Surg Oral Med Oral Pathol* 1986;61:373-81.
 27. Bosman FT. Dysplasia classification: Pathology in disgrace? *J Pathol* 2001;194:143-4.